GENES ASSOCIATED WITH NEUROTRANSMITTER PROCESSING

TECHNICAL FIELD

The invention relates to five genes associated with neurotransmitter processing and their corresponding gene products. The invention also relates to the use of these biomolecules in diagnosis, treatment, prevention, and evaluation of therapies for neurological and hormone-related diseases, more particularly Parkinson's disease, schizophrenia, and epilepsy.

BACKGROUND OF THE INVENTION

Normal neurological function requires a careful balance of neurotransmitters. An upset in the balance can lead to disease. For example, Parkinson's disease is triggered by an as yet unknown event that increases the susceptibility of dopaminergic neurons to oxidative processes which in turn leads to cell damage and death (Jenner, P. (1998) Mov. Disord. 13:24-34). The resulting lack of dopamine leads to tremors and rigidity in the patient, the main clinical manifestations of Parkinson's disease (Birkmayer, W. and P. Riederer (1986) Understanding the Neurotranmitters: Keys to the Workings of the Brain, Springer-Verlag, New York NY).

Many other disorders result from an improper balance among neurotransmitters. The central dopaminergic neurons have long been thought to play a significant role in schizophrenia (Farde, L. (1997) Schizophr. Res. 28:157-162), but a more complex picture is emerging where interactions among dopamine, norepinephrine and other neurotransmitters appears to be the basis of the disorder (Carlsson, A. et al. (1997) Life Sci. 61:75-94). Both dopamine and norepinpehrine are also implicated in attention deficit hyperactive disorder (Hechtman, L. (1994) J. Psychiary Neurosci. 19: 193-201). In animal models norepinephrine has been implicated in epilepsy (Szot, P. et al. (1996) Brain Res. Mol. Brain Res. 43:233-245; Clough, R.W. et al. (1998) Epilepsy Res. 29:135-146); Janumpalli, S. et al. (1998) J. Neurosci. 18:2004-2008). Norepinephrine directly regulates gonadotropin-releasing hormone and may therefore be involved in reproductive disorders (Becu-Villalobos, D. et al. (1997) Cell Mol. Neurobiol. 17:699-715; Pau, K.Y. and H.G. Spies (1997) Chin. J. Physiol. 40:181-196). Norepinephrine also regulates growth hormone and inhibition of dopamine β hydroxylase (DBH). DBH retards growth in animal models (Malozowski, S. et al. (1993) Acta Endocrinol. 129:554-558).

In general, neurological disorders are treated with drugs that modulate neurotransmitter

level in an attempt to maintain the correct level. Consequently, any of the substrates, enzymes, receptors or neurotransmitter transporters as well as genes involved in their regulation are potentially of therapeutic use. We have discovered five new genes that exhibit significant coexpression with genes known to be involved in the synthesis and release of the

5   neurotransmitters dopamine, norepinephrine and epinephrine. These new genes are useful in the diagnosis, prevention, treatment, and evaluation of therapies for diseases, particularly neurological and hormone-related diseases, more particularly Parkinson's disease, schizophrenia, and epilepsy.

SUMMARY OF THE INVENTION

In one aspect, the invention provides for a substantially purified polynucleotide

10   comprising a gene that is coexpressed with one or more neurotransmitter-processing-specific genes in a plurality of biological samples. Preferably, each neurotransmitter-processing-specific gene is selected from (a) a polynucleotide sequence selected from the group consisting of SEQ ID NOs: 1-5; (b) a polynucleotide sequence which encodes the polypeptide sequence of SEQ ID NO: 6; (c) a polynucleotide sequence which is complementary to the polynucleotide sequence of (a)

15   or (b); and (d) a probe which hybridizes to the polynucleotide of (a), (b), or (c). Furthermore, the invention provides an expression vector comprising any of the above described polynucleotides and host cells comprising the expression vector.

In a second aspect, the invention provides a substantially purified polypeptide comprising the gene product of a gene that is coexpressed with one or more neurotransmitter-processing-

20   specific genes in a plurality of biological samples. The neurotransmitter-processing-specific gene may be selected from the group consisting of L-tyrosine hydroxylase (TH), aromatic amino acid decarboxylase (AADC), dopamine β-hydroxylase (DBH), nicotinic acetylcholine receptor α3 subunit precursor (nAchR-α3), secretogranin I and II, Rab3a, human cocaine and amphetamine regulated transcript (hCART), vesicular monoamine transporter 1 (hVMAT1), and ARIX

25   homeodomain protein. Preferred embodiments are (a) the polypeptide sequence of SEQ ID NO: 6, and (b) a polypeptide sequence comprising at least 6 sequential amino acids of the polypeptide sequence of (a) . Additionally, the invention provides antibodies that bind specifically to any of the above described polypeptides.

In another aspect, the invention provides a pharmaceutical composition comprising the

30   polynucleotide or the polypeptide of the inventions in conjunction with a suitable pharmaceutical carrier and a method for treating or preventing disease comprising administering to a subject in need such a composition in an amount effective for treating or preventing said disease.

In yet a further aspect, the invention provides a method for diagnosing a disease or condition associated with the altered expression of a gene that is coexpressed with one or more

neurotransmitter-processing-specific genes in a sample, wherein each neurotransmitter-processing-specific gene is selected from the group consisting of L-tyrosine hydroxylase (TH), aromatic amino acid decarboxylase (AADC), dopamine β-hydroxylase (DBH), nicotinic acetylcholine receptor α3 subunit precursor (nAchR-α3), secretogranin I and II, Rab3a, human

5   cocaine and amphetamine regulated transcript (hCART), vesicular monoamine transporter 1 (hVMAT1), and ARIX homeodomain protein. The method comprises providing the sample comprising one of more of said coexpressed genes; hybridizing the polynucleotide to said coexpressed genes under conditions effective to form one or more hybridization complexes; and detecting the hybridization complexes, wherein the presence of one or more of the hybridization

10   complexes correlates with the presence of the disease or condition.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

The Sequence Listing provides exemplary gene sequences associated with neurotransmitter processing, SEQ ID NOs: 1-5, and a polypeptide sequence, SEQ ID NO: 6, derived from the gene sequence of SEQ ID NO:4. Each sequence is identified by a numerical

15   sequence identification number and by the Incyte clone number from which the sequence was first identified.

DESCRIPTION OF THE INVENTION

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include the plural reference unless the context clearly dictates otherwise. Thus,

20   for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

DEFINITIONS

"NSEQ" refers generally to a polynucleotide sequence of the present invention, including

25   SEQ ID NOs: 1-5. "PSEQ" refers generally to a polypeptide sequence of the present invention, including SEQ ID NO: 6.

A " variant" refers to either a polynucleotide or a polypeptide whose sequence diverges from SEQ ID NOs: 1-5 or SEQ ID NO: 6, respectively. Polynucleotide sequence divergence may result from mutational changes such as deletions, additions, and substitutions of one or more

30   nucleotides; it may also occur because of differences in codon usage. Each of these types of changes may occur alone, or in combination, one or more times in a given sequence. Polypeptide variants include sequences that possess at least one structural or functional characteristic of SEQ ID NO: 6.

A "fragment" can refer to a nucleic acid sequence that is preferably at least 20 nucleic

acids in length, more preferably 40 nucleic acids, and most preferably 60 nucleic acids in length, and encompasses, for example, fragments consisting of nucleic acids 1-50 of SEQ ID NOs:1-5. A "fragment" can also refer to polypeptide sequences which are preferably at least 5 to about 15 amino acids in length, most preferably at least 10 amino acids long, and which retain some

5   biological activity or immunological activity of, for example, SEQ ID NO:6.

"Neurotransmitter-processing-specific genes" refers to genes which are involved in the processing, such as the synthesis or release, of neurotransmitters. For purposes of this invention, the neurotransmitters are preferably dopamine, norepinephrine and epinephrine.

"Gene associated with neurotransmitter processing" refers to a new gene sequence whose

10   expression pattern is similar to that of the neurotransmitter-processing-specific genes and which is useful in the diagnosis, treatment, or prevention of diseases associated with the altered expression of neurotransmitter processing genes. The gene sequences can also be used in the evaluation of therapies for these diseases.

"Substantially purified" refers to a nucleic acid or an amino acid sequence that is removed

15   from its natural environment and is isolated or separated, and is at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which it is naturally present.

THE INVENTION

The present invention encompasses five genes useful in the diagnosis, treatment,

20   prevention, and evaluation of therapies for diseases related to improper neurotransmitter processing, particularly neurological and growth-related diseases, more particularly Parkinson's disease, schizophrenia, and epilepsy. These genes were identified using the method disclosed in the US patent application entitled "Prostate Cancer-Associated Genes", Walker et al., Serial No:09/102,615 filed June 22, 1998, herein incorporated by reference, and described briefly below.

25   The method provides first identifying polynucleotides that are expressed in a plurality of cDNA libraries. The polynucleotides may include genes of known function, genes known to be specifically expressed in a specific disease process, subcellular compartment, cell type, tissue type, or species. Additionally, the polynucleotides include genes of unknown function. The expression patterns of the known genes are then compared with those of the genes of unknown function to

30   determine whether a specified coexpression probability threshold is met. Through this comparison, a subset of the polynucleotides having a high coexpression probability with the known genes can be identified. The high coexpression probability correlates with a particular coexpression probability threshold which is less than 0.001, and more preferably less than 0.00001.

The polynucleotides may originate from cDNA libraries derived from a variety of sources including, but not limited to, eukaryotes such as human, mouse, rat, dog, monkey, plant, and yeast and prokaryotes such as bacteria and viruses. These polynucleotides can also be selected from a variety of sequence types including, but not limited to, expressed sequence tags (ESTs), assembled

5   polynucleotide sequences, full length gene coding regions, introns, regulatory sequences, 5' untranslated regions, and 3' untranslated regions. To have statistically significant analytical results, the polynucleotides need to be expressed in at least three cDNA libraries. In one preferred embodiment, the polynucleotides are obtained from sequence databases, such as the LIFESEQ database (Incyte Pharmaceuticals, Palo Alto CA).

10   The cDNA libraries used in the coexpression analysis of the present invention may be obtained from blood vessels, heart, blood cells, cultured cells, connective tissue, epithelium, islets of Langerhans, neurons, phagocytes, biliary tract, esophagus, gastrointestinal system, liver, pancreas, fetus, placenta, chromaffin system, endocrine glands, ovary, uterus, penis, prostate, seminal vesicles, testis, bone marrow, immune system, cartilage, muscles, skeleton, central

15   nervous system, ganglia, neuroglia, neurosecretory system, peripheral nervous system, bronchus, larynx, lung, nose, pleurus, ear, eye, mouth, pharynx, exocrine glands, bladder, kidney, ureter, and the like. The number of cDNA libraries selected can range from as few as 20 to greater than 10,000. Preferably, the number of the cDNA libraries is greater than 500.

In a preferred embodiment, gene sequences are assembled to reflect related sequences,

20   such as assembled sequence fragments derived from a single transcript. Assembly of the polynucleotide sequences can be performed using sequences of various types including, but not limited to, ESTs, extensions, or shotgun sequences. In a most preferred embodiment, the polynucleotide sequences are derived from human sequences that have been assembled using the algorithm disclosed in the US provisional patent application entitled "Database and System for

25   Storing, Comparing and Displaying Related Biomolecular Sequence Information", Lincoln et al., Serial No:60/079,469, filed March 26, 1998, herein incorporated by reference.

Experimentally, differential expression of the polynucleotides can be evaluated by methods including, but not limited to, differential display by spatial immobilization or by gel electrophoresis, genome mismatch scanning, representational difference analysis, and transcript

30   imaging. Additionally, differential expression can be assessed by microarray technology. These methods may be used alone or in combination.

Neurotransmitter-processing-specific genes can be selected based on the current use of these genes as diagnostic markers or as therapeutic targets for diseases related to incorrect neurotransmitter processing, such as the neurotransmitter processing. Genes related to

neurotransmitter processing include L-tyrosine hydroxylase (TH), aromatic amino acid decarboxylase (AADC), dopamine β-hydroxylase (DBH), nicotinic acetylcholine receptor α3 subunit precursor (nAchR-α3), secretogranin I and II, Rab3a, human cocaine and amphetamine regulated transcript (hCART), vesicular monoamine transporter 1 (hVMAT1), ARIX

5   homeodomain protein, and the like.

The procedure for identifying new genes that exhibit a statistically significant coexpression pattern with known genes related to neurotransmitter processing is as follows. First, the presence or absence of a gene sequence in a cDNA library is defined: a gene is present in a cDNA library when at least one cDNA fragment corresponding to that gene is detected in a cDNA

10   sample taken from the library, and a gene is absent from a library when no corresponding cDNA fragment is detected in the sample.

Second, the significance of gene coexpression is evaluated using a probability method to measure a due-to-chance probability of the coexpression. The probability method can be the Fisher exact test, the chi-squared test, or the kappa test. These tests and examples of their

15   applications are well known in the art and can be found in standard statistics texts (Agresti, A. (1990 Categorical Data Analysis. Wiley, New York NY; Rice, J. A. (1988) Mathematical Statistics and Data Analysis. Wadsworth & Brooks/Cole, Pacific Grove CA). A Bonferroni correction (Rice, supra, page 384) can also be applied in combination with one of the probability methods for correcting statistical results of one gene versus multiple other genes. In a preferred embodiment,

20   the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set to less than 0.001, more preferably less than 0.00001.

To determine whether two genes, A and B, have similar coexpression patterns, occurrence data vectors can be generated as illustrated in Table 1, wherein a gene's presence is indicated by a one and its absence by a zero. A zero indicates that the gene did not occur in the library, and a

25   one indicates that it occurred at least once.

Table 1. Occurrence data for genes A and B

	Library 1	Library 2	Library 3	...	Library N
gene A	1	1	0	...	0
gene B	1	0	1	...	0

30

For a given pair of genes, the occurrence data in Table 1 can be summarized in a 2x2 contingency table.

Table 2. Contingency table for co-occurrences of genes A and B

	Gene A present	Gene A absent	Total
Gene B present	8	2	10
Gene B absent	2	18	20
Total	10	20	30

Table 2 presents coexpression data for gene A and gene B in a total of 30 libraries. Gene A and gene B each occurs 10 times in the libraries. Table 2 summarizes and presents 1) the number of times gene A and B are both present in a library, 2) the number of times gene A and B are both absent in a library, 3) the number of times gene A is present while gene B is absent, and 4) the number of times gene B is present while gene A is absent. The upper left entry is the number of times the two genes co-occur in a library, and the middle right entry is the number of times neither gene occurs in a library. The off diagonal entries are the number of times one gene occurs while the other does not. Gene A and B are both present eight times and absent 18 times, gene A is present while gene B is absent two times, and gene B is present while gene A is absent two times. The probability ("p-value") that the above association occurs due to chance as calculated using a Fisher Exact test is 0.0003. Associations are generally considered significant if a p-value is less than 0.01 (Agresti, supra; Rice, supra).

This method of estimating the probability for coexpression of two genes makes several assumptions. The method assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent because more than one library may be obtained from a single patient or tissue, and they are not entirely identically sampled because different numbers of cDNA's may be sequenced from each library (typically ranging from 5,000 to 10,000 cDNAs per library). In addition, because a Fisher exact coexpression probability is calculated for each gene versus 41,419 other genes, the probability is corrected for the multiple statistical tests by requiring a more stringent p-value than the standard p-value of 0.01.

Using the coexpression analysis method, we have identified five new genes that exhibit strong association, or coexpression, with known genes related to neurotransmitter processing. The results presented in Tables 5 to 10 show that the expression of the five new genes have direct or indirect association with the expression of known genes related to neurotransmitter processing. Therefore, the new genes can potentially be used in diagnosis, treatment, or prevention of diseases related to neurotransmitter processing, such as Parkinson's disease, schizophrenia, epilepsy, female reproductive disorders, attention deficit disorder, or in the evaluation of therapies for these

diseases. Further, the gene products of the five new genes are potential therapeutic proteins and targets of therapeutics against these diseases.

Therefore, in one embodiment, the present invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NOs: 1-5. These five polynucleotides are shown by the method of the present invention to have strong coexpression association with known genes related to neurotransmitter processing. The invention also encompasses a variant of the polynucleotide sequence, its complement, or 18 consecutive nucleotides of the sequences provided in the above described sequences. Variant polynucleotide sequences typically have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to NSEQ.

One preferred method for identifying variants provides using NSEQ and/or PSEQ sequences to search against the GenBank primate (pri), rodent (rod), and mammalian (mam), vertebrate (vrtp), and eukaryote (eukp) databases, SwissProt, BLOCKS (Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221), PFAM, and other databases that contain previously identified and annotated motifs, sequences, and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith, T. et al. (1992) Protein Engineering 5:35-51) as well as algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410), BLOCKS (Henikoff S. and Henikoff G.J. (1991) Nucleic Acids Research 19:6565-6572), Hidden Markov Models (HMM; Eddy, S.R. (1996) Cur. Opin. Str. Biol. 6:361-365; and Sonnhammer, E.L.L. et al. (1997) Proteins 28:405-420), and the like, can be used to manipulate and analyze nucleotide and amino acid sequences. These databases, algorithms and other methods are well known in the art and are described in Ausubel, F.M. et al. (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY) and in Meyers, R.A. (1995; Molecular Biology and Biotechnology, Wiley VCH, Inc, New York NY, p 856-853).

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to SEQ ID NOs: 1-5, and fragments thereof under stringent conditions. Stringent conditions can be defined by salt concentration, temperature, and other chemicals and conditions well known in the art.

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional

-8-

parameters, such as hybridization time, the concentration of detergent (sodium dodecyl sulfate, SDS) or solvent (formamide), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Additional variations on these conditions will be readily apparent to those skilled in the art (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel,

5   A.R. (1987) Methods Enzymol. 152:507-511; Ausubel, supra; and Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

NSEQ or the polynucleotide sequences encoding PSEQ can be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to obtain full length sequences or to detect upstream sequences, such as promoters and regulatory elements.

10   (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995; PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5; Sarkar, G. (1993; PCR Methods Applic. 2:318-322); Triglia, T. et al. (1988; Nucleic Acids Res. 16:8186); Lagerstrom, M. et al. (1991; PCR Methods Applic. 1:111-119); and Parker, J.D. et al. (1991; Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR and nested primers to walk genomic DNA. This procedure avoids the need to

15   screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 software (National Biosciences, Plymouth MN) or another appropriate program, to be about 18 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

20   In another aspect of the invention, NSEQ or the polynucleotide sequences encoding PSEQ can be cloned in recombinant DNA molecules that direct expression of PSEQ or the polypeptides encoded by NSEQ, or structural or functional fragments thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express the

25   polypeptides of PSEQ or the polypeptides encoded by NSEQ. The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter the nucletide sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used

30   to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In order to express a biologically active polypeptide encoded by NSEQ, NSEQ or the polynucleotide sequences encoding PSEQ, or derivatives thereof, may be inserted into an

appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in NSEQ or polynucleotide sequences

5   encoding PSEQ. Methods which are well known to those skilled in the art may be used to construct expression vectors containing NSEQ or polynucleotide sequences encoding PSEQ and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, supra, and Ausubel, supra).

10   A variety of expression vector/host cell systems may be utilized to contain and express NSEQ or polynucleotide sequences encoding PSEQ. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (baculovirus); plant cell systems transformed with viral

15   expression vectors, cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV), or with bacterial expression vectors (Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed. For long term production of recombinant proteins in mammalian systems, stable expression of a polypeptide encoded by NSEQ in cell lines is preferred. For example, NSEQ or sequences encoding PSEQ can be transformed into cell lines

20   using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector.

          In general, host cells that contain NSEQ and that express PSEQ may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or

25   immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences. Immunological methods for detecting and measuring the expression of PSEQ using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell

30   sorting (FACS).

          Host cells transformed with NSEQ or polynucleotide sequences encoding PSEQ may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art,

-10-

expression vectors containing polynucleotides of NSEQ or polynucleotides encoding PSEQ may be designed to contain signal sequences which direct secretion of PSEQ or polypeptides encoded by NSEQ through a prokaryotic or eukaryotic cell membrane.

5   In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for
10   post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

       In another embodiment of the invention, natural, modified, or recombinant NSEQ or nucleic acid sequences encoding PSEQ are ligated to a heterologous sequence resulting in
15   translation of a fusion protein containing heterologous protein moieties in any of the aforementioned host systems. Such heterologous protein moieties facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, hemagglutinin (HA) and monoclonal antibody
20   epitopes..

       In another embodiment, NSEQ or sequences encoding PSEQ are synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. (7):215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. (7):225-232; and Ausubel, supra). Alternatively, PSEQ or a polypeptide sequence encoded by
25   NSEQ itself, or a fragment thereof, may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A Peptide synthesizer (PE Biosystems, Foster City CA). Additionally, PSEQ or the amino acid sequence encoded by NSEQ, or any part thereof, may be altered during direct synthesis and/or combined
30   with sequences from other proteins, or any part thereof, to produce a polypeptide variant.

       In another embodiment, the invention provides a substantially purified polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 6, or fragments thereof.

DIAGNOSTICS and THERAPEUTICS

The genes of the invention can be used in diagnosis, prevention, treatment, and evaluation of therapies for diseases related to neurotransmitter processing, particularly neurological and hormone-related diseases, more particularly Parkinson's disease, schizophrenia, epilepsy, female
5   reproductive disorders, and attention deficit disorder. Further, the amino acid sequences encoded by the new genes are potential therapeutic proteins and targets of therapeutics against these diseases.

In one preferred embodiment, the NSEQ or the polynucleotides encoding PSEQ are used for diagnostic purposes to determine the absence, presence, and excess expression of PSEQ, and
10   to monitor regulation of the levels of mRNA or the polypeptides encoded by NSEQ during therapeutic intervention. The polynucleotides may be at least 18 nucleotides long, complementary RNA and DNA molecules, branched nucleic acids, and peptide nucleic acids (PNAs). Alternatively, the polynucleotides are used to detect and quantitate gene expression in samples in which expression of PSEQ or the polypeptides encoded by NSEQ are correlated with disease.
15   Additionally, NSEQ or the polynucleotides encoding PSEQ can be used to detect genetic polymorphisms associated with a disease. These polymorphisms may be detected at the transcript cDNA or genomic level.

The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the
20   hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding PSEQ, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the NSEQ or PSEQ-encoding sequences.
25   Means for producing specific hybridization probes for DNAs encoding PSEQ include the cloning of NSEQ or polynucleotide sequences encoding PSEQ into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter
30   groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, by fluorescent labels and the like. The polynucleotide sequences encoding PSEQ may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies;and in microarrays utilizing fluids or tissues from patients to detect altered PSEQ expression. Such qualitative or quantitative

-12-

methods are well known in the art.

NSEQ or the nucleotide sequences encoding PSEQ can be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal

5   is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to the standard value then the presence of altered levels of nucleotide sequences of NSEQ and those encoding PSEQ in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an

10   individual patient.

Once the presence of a disease is established and a treatment protocol is initiated, hybridization or amplification assays can be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a

15   period ranging from several days to months.

The polynucleotides may be used for the diagnosis of a variety of diseases associated with neurotransmitter processing such as neurological disorders including, but not limited to, akathesia, Alzheimer's disease, amnesia, amylotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia,

20   dystonias, epilepsy, Huntington's disease, peripheral neuropathy, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherapeutic neuralgia, schizophrenia, and Tourette's disorder; and reproductive disorders including, but not limited to, disorders of prolactin production, tubal disease, ovulatory defects, endometriosis, disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian

25   hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, testis, and prostate, fibrocystic breast disease, galactorrhea disruptions of spermatogenesis, abnormal sperm physiology, prostatitis, Peyronie's disease, and gynecomastia.

Alternatively, the polynucleotides may be used as targets in a microarray. The microarray

30   can be used to monitor the expression level of large numbers of genes simultaneously and to identify splice variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disease, to diagnose a disease, and to develop and monitor the activities of therapeutic agents.

In yet another alternative, polynucleotides may be used to generate hybridization probes

useful in mapping the naturally occurring genomic sequence. Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968).

In another embodiment, antibodies which specifically bind PSEQ may be used for the diagnosis of diseases characterized by the over- or under-expression of PSEQ or polypeptides encoded by NSEQ. Alternatively, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PSEQ or the polypeptides encoded by NSEQ specifically compete with a test compound for binding the polypeptides. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PSEQ or the polypeptides encoded by NSEQ. Diagnostic assays for PSEQ or the polypeptides encoded by NSEQ include methods which utilize the antibody and a label to detect PSEQ or the polypeptided encoded by NSEQ in human body fluids or in extracts of cells or tissues. A variety of protocols for measuring PSEQ or the polypeptides encoded by NSEQ, including ELISAs, RIAs, and FACS, are well known in the art and provide a basis for diagnosing altered or abnormal levels of the expression of PSEQ or the polypeptides encoded by NSEQ. Normal or standard values for PSEQ expression are established by combining body fluids or cell extracts taken from normal subjects, preferably human, with antibody to PSEQ or a polypeptide encoded by NSEQ under conditions suitable for complex formation The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of PSEQ or the polypeptides encoded by NSEQ expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing or monitoring disease.

In another aspect, the polynucleotides and polypeptides of the present invention can be employed for treatment or the monitoring of therapeutic treatments for the diseases specified above. The polynucleotides of NSEQ or those encoding PSEQ, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotides of NSEQ or those encoding PSEQ may be used in situations in which it would be desirable to block the transcription or translation of the mRNA.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding PSEQ. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes having polynucleotide sequences of NSEQ or those encoding PSEQ can be turned

off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding PSEQ. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are

5   preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, <u>Molecular and Immunologic Approaches</u>, Futura

10   Publishing Co., Mt. Kisco NY, pp. 163-177.) Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase

15   linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

20   Many methods for introducing vectors into cells or tissues are available and equally suitable for use <u>in vivo</u>, <u>in vitro</u>, and <u>ex vivo</u>. For <u>ex vivo</u> therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al.

25   (1997) Nature Biotechnology 15:462-466.)

Further, an antagonist or antibody of a polypeptide of PSEQ or encoded by NSEQ may be administered to a subject to treat or prevent a disease related to synthesis and release of dopamine and norepinephrine with increased expression or activity of PSEQ. An antibody which specifically binds the polypeptide may be used directly as an antagonist or indirectly as a targeting

30   or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the polypeptide.

Antibodies to PSEQ or a polypeptide encoded by NSEQ may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments

produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use. Monoclonal antibodies to PSEQ may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the

5   human B-cell hybridoma technique, and the EBV-hybridoma technique. In addition, techniques developed for the production of chimeric antibodies can be used. (See, e.g., Meyers, supra.) Alternatively, techniques described for the production of single chain antibodies may be employed. Antibody fragments which contain specific binding sites for PSEQ or the polypeptide sequences encoded by NSEQ may also be generated.

10   Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art.

Yet further, an agonist of a polypeptide of PSEQ or that encoded by NSEQ may be

15   administered to a subject to treat or prevent a cancer associated with decreased expression or activity of the polypeptide.

An additional aspect of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of

20   polypeptides of PSEQ or those encoded by NSEQ, antibodies to the polypeptides, and mimetics, agonists, antagonists, or inhibitors of the polypeptides. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient

25   alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

30   In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co. Easton PA).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example, polypeptides of PSEQ or those encoded by NSEQ, or fragments thereof, antibodies of the polypeptides, and agonists, antagonists or inhibitors of the polypeptides, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

EXAMPLES

It is understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary. It is also understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. The examples below are provide to illustrate the subject invention and are not included for the purpose of limiting the invention.

I. PGANNOT01 cDNA Library Construction

For purposes of example, the preparation of the PGANNOT01 library is described. The PGANNOT01 cDNA library was constructed using polyA RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, a myocardial infarction, and type II diabetes.

The frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a Polytron homogenizer (PT-3000; Brinkmann Instruments, Westbury NJ). The lysate was centrifuged over a 5.7 M CsCl cushion using an SW28 rotor in a BL8-70M ultracentrifuge (Beckman Coulter, Fullerton CA) for 18 hours at 25,000 rpm at ambient temperature. The RNA

was extracted twice with acid phenol, pH 4.0, following an RNA isolation protocol (Stratagene, La Jolla CA), precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in DEPC-treated water and treated with DNase for 15 min at 37°C. The reaction was stopped with an equal volume of acid phenol and the RNA was isolated using the OLIGOTEX kit (Qiagen, Chatsworth CA) and used to construct the cDNA library.

The RNA was handled according to the recommended protocols in the SUPERSCRIPT Plasmid system (Life Technologies, Gaithersburg MD), and cDNAs were ligated into pSport I plasmid (Life Technologies). The plasmid was subsequently transformed into DH5α competent cells (Life Technologies).

II   Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the MINIPREP kit (Edge Biosystems, Gaithersburg MD). This kit consists of a 96 well block with reagents for 960 purifications. The recommended protocol was employed except for the following changes: 1) the 96 wells were each filled with only 1 ml of sterile Terrific Broth (Life Technologies) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) the bacteria were cultured for 24 hours after the wells were inoculated and then lysed with 60 μl of lysis buffer; 3) a centrifugation step employing the GS-6R (Beckman Coulter) at 2900 rpm for 5 min was performed before the contents of the block were added to the primary filter plate; and 4) the optional step of adding isopropanol to TRIS buffer was not routinely performed. After the last step in the protocol, samples were transferred to a 96-well block for storage.

Alternative methods of purifying plasmid DNA include the use of MAGIC MINIPREPS system (Promega, Madison WI) or QIAWELL-8 Plasmid, QIAWELL PLUS DNA and QIAWELL ULTRA DNA purification systems (Qiagen).

The cDNAs were prepared using a MICROLAB 2200 system (Hamilton, Reno NV) in combination with four DNA ENGINE thermal cyclers (PTC200; MJ Research, Watertown MA) and sequenced by the method of Sanger F and AR Coulson (1975; J Mol Biol 94:441f) using ABI PRISM 377 or 373 DNA sequencing systems (PE Biosystems).

III.   Selection, Assembly, and Characterization of Sequences

The sequences used for coexpression analysis were assembled from EST sequences, 5' and 3' longread sequences, and full length coding sequences. Selected sequences were expressed in at least three cDNA libraries.

The assembly process is described as follows. EST sequence chromatograms were processed and verified. Quality scores were obtained using PHRED (Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194). The edited

sequences were loaded into a relational database management system (RDBMS). The EST sequences were clustered into an initial set of bins using BLAST with a product score of 50. All clusters of two or more sequences were created as bins. The overlapping sequences represented in a bin correspond to the sequence of a transcribed gene.

5   Assembly of the component sequences within each bin was performed using a modification of Phrap, a publicly available program for assembling DNA fragments (Green, P. University of Washington, Seattle WA). Bins that showed 82% identity from a local pair-wise alignment between any of the consensus sequences were merged.

        Bins were annotated by screening the consensus sequence in each bin against public
10   databases, such as GBPRI and GENPEPT from NCBI. The annotation process involved a FASTn screen against the GBPRI database in GENBANK. Those hits with a percent identity of greater than or equal to 70% and an alignment length of greater than or equal to 100 base pairs were recorded as homolog hits. The residual unannotated sequences were screened by FASTx against GENPEPT. Those hits with an E value of less than or equal to 10^{-8} are recorded as homolog hits.

15   Sequences were then reclustered using BLASTn and Cross-Match, a program for rapid protein and nucleic acid sequence comparison and database search (Green, P. University of Washington, Seattle WA), sequentially. Any BLAST alignment between a sequence and a consensus sequence with a score greater than 150 was realigned using cross-match. The sequence was added to the bin whose consensus sequence gave the highest Smith-Waterman score amongst
20   local alignments with at least 82% identity. Non-matching sequences created new bins. The assembly and consensus generation processes were performed for the new bins.

IV.   Known Genes Used for Identifying New Genes Associated with Neurotransmitter Processing

        Any of ten neurotransmitter-processing-specific genes were used to identify new Incyte
25   genes, using the coexpression analysis method described above. The genes included L-tyrosine hydroxylase (TH), aromatic amino acid decarboxylase (AADC), dopamine β hydroxylase (DBH), nicotinic acetylcholine receptor α3 subunit precursor (nAchR-α3), secretogranin I and II, Rab3a, human cocaine and amphetamine regulated transcript (hCART), vesicular monoamine transporter 1 (hVMAT1), ARIX homeodomain protein, and the like. Other known genes that have potential
30   association with dopamine and norepinephrine pathways include endothelin converting enzymes (ECEs), rap2-interacting protein 8 (RPIP8), neuron-specific growth-associated protein (SCG10), and "Delta-like" putative homeotic protein (dlk).

        L-tyrosine hydroxylase (TH) catalyzes the conversion of L-tyrosine to L-dopa, which is the first and rate-limiting step in the synthesis of the catecholamines dopamine, norepinephrine and

epinephrine (Nagatsu, T. (1995) Essays Biochem. 30:15-35). TH activity is regulated in several ways including transcriptionally, by alternative mRNA splicing, and by regulation of mRNA stability. Aromatic amino acid decarboxylase (AADC) catalyzes the decarboxylation of L-tyorsine, L-tryptophan, L-phenylalanine and L-dopa. Decarboxylation of L-dopa yields dopamine (Webster

5   and Jordan, supra), and increased levels of AADC enhanced dopamine synthesis in vivo (Kaddis et al., supra). Dopamine β hydroxylase (DBH) converts dopamine to norepinephrine (Wu, H.J. et al. (1990) J. Neurochem. 55:97-105). Nicotinic acetylcholine receptor α3 subunit precursor (nAchR-α3) is a subunit in the ligand-gated sodium ion channel triggered by nicotine and acetylcholine. Prolonged stimulation of nAchR with nicotine or acetylcholine induces TH and DBH (Patrick, R.L.

10   and J.D. Barchas (1976) J. Pharmacol. Exp. Ther. 197:181-196). Secretogranins are vesicle associated proteins. Secretogranin I is found in norepinephrine-containing vesicles (Bressler, J.P. et al. (1996) J. Neurosci. Res. 46:678-685). Secretogranin II is a precursor for secretoneurin, which stimulates dopamine release (Agneter, E. et al. (1995) J. Neurochem. 65:622-625; Fischer-Colbrie, R. et al. (1995) Prog. Neurobiol. 46:49-70). Secretogranin II occurs in dopamine- and

15   norepinephrine-containing vesicles in neurons (Goodall, A.R. et al. (1997) J. Neurochem. 68:1542-1552). Rab3a is a member of the Ras GTPase family. Rab3a is associated with large dense core vesicles and partially co-localizes with DBH (Johannes, L. et al. (1994) Embo. J. 13:2029-2037; Darchen, F. et al. (1995) J. Cell Sci. 108: 1639-1649), and is one of the factors controlling Ca2+-dependent exocytosis (Johannes et al., supra). Human cocaine and amphetamine regulated

20   transcript (hCART) is a signal peptide protein and is specifically induced in neurons exposed to cocaine and amphetamine (Douglass, J. and S. Daoud (1996) Gene 169:241-245). Cocaine blocks re-uptake of dopamine from the synapse (Reith, M.E. et al. (1997) Eur. J. Pharmacol. 324:1-10 as does amphetamine, which also induces dopamine release (Jones, S.R. et al. (1998) J. Neurosci. 18:1979-1986). Vesicular monoamine transporter 1 (hVMAT1) packages the neurotransmitter

25   monoamines into large dense core vesicles for exocytosis at the synapse (Liu, Y. et al. (1994) J. Cell Biol. 127:1419-1433). ARIX is a human homeodomain protein which is specifically expressed in noradrenergic, DBH-positive tissues and in cell lines derived from those tissues (Zellmer, E. et al. (1995) J. Neurosci. 15:8109-8120). ARIX regulates the transcriptional activities of the DBH promoter and the promoter of the tyrosine hydroxylase gene, two of the promoters

30   which control the specific expression of catecholamine biosynthetic genes.

The endothelins, ET-1, ET-2, and ET-3, are a class of secreted peptides that are processed from inactive precursor peptides. The third and final processing step leading to the active ET is catalyzed by a member of the endothelin converting enzymes (ECEs). The ET's were known vasoactive agents, potential neurotransmitters, and potential growth factors (Barnes, K. and A.J.

Turner (1997) Neurochem. Res. 22:1033-1040). ET-1 was shown to enhance norepinephrine pressor response (Kita, S. et al. (1998) J. Cardiovasc. Pharmacol. 31:S119-S121), and ET-1 and ET-3 to induce dopamine release (Horie, K. et al. (1995) Life Sci. 57:735-741; Shibaguchi, H. et al. (1997) Cell Mol. Neurobiol. 17:471-481). ET-2, however, has not been shown in published

5 literature to be associated with norepinephrine or dopamine processing. Rap2-interacting protein 8 (RPIP8) was discovered using a yeast two-hybrid screen of mouse brain cDNAs using the small GTP-binding protein Rap2 as bait (Janoueix-Lerosey, I. et al. (1998) Eur. J. Biochem. 252(2):290-298). Rap2 is a member of the Ras family, another member of which (Rab3a) was known to be involved in exocytosis of neurotransmitters. SCG10 is a neuron-specific growth-associated protein

10 which belongs to the stathmin family. SCG10 functions in increasing the breakdown of microtubules in the neural growth cone (Riederer, B. M. et al. (1997) Proc. Natl. Acad. Sci. 91:741-745). dlk is a "Delta-like" putative homeotic protein thought to be involved in neuroendocrine differentiation (Laborda, J. et al. (1993) J. Biol. Chem. 268:3817-3820). It belongs to the epidermal growth factor receptor family (Lee, Y.L. et al. (1995) Biochim. Biophys.

15 Acta 1261:223-232) and is induced by growth hormone (Carlsson, C. et al. (1997) Endocrinology 138:3940-3948) and regulated through ErbB3 (Edman, C.F. et al. (1997) Biochem. J. 323:113-118).

Table 4 shows the five genes that have the strongest association with TH, a gene involved in the synthesis and release of dopamine, and norepinephrine. These coexpressed genes are

20 presented with their p-values. The column headings have the following meanings:

P-value	The probability that the observed number of co-occurrences is due to chance using the Fisher Exact method.
Associated gene	A gene that shows significant co-expression with the target.
Occurs	The number of libraries in which the associated gene occurs.
Both occur	The number of libraries in which both the target gene and the co-expressed gene occur.
Target only	The number of libraries in which only the target gene occurs.
Associated only	The number of libraries in which only the associated gene occurs.
Neither	The number of libraries in which neither the target gene nor the associated gene occur.

Table 4 Co-expression results for TH

P-value	Associated gene	Occurs	Both occur	Target only	Associated only	Neither
1.47E-13	DBH	12	9	6	3	503
6.11E-11	nAchR-α3	13	8	7	5	501

6.22E-10	Secretogranin 1	45	11	4	34	472
3.51E-09	Human dlk mRNA	52	11	4	41	465
1.01E-08	Homolog to ECE	13	7	8	6	500

5   As a target, TH occurred in 15 of 522 cDNA libraries studied, and showed strong coexpression with genes known to be related to neurotransmitter processing and other known genes, such as DBH, nAchR, secretogranin I, human dlk mRNA for putative homeotic protein, and a homolog of endothelin converting enzyme (ECE). These results are shown in Table 4 with association probability in the range of 1.58E-8 to 1.53E-13.

10   Similar results were observed when the other nine known genes, AADC, DBH, nAchR-α3, secretogranin I and II, Rab3a, hCART, hVMAT1, and ARIX were taken as target genes. These results indicate that the coexpression method was successful in identifying association of the known genes among themselves and with the known genes specified above, and thus was effective in identifying genes that are associated with the synthesis and release of neurotransmitters such as

15   dopamine and norepinephrine.

V.   Identification of New Genes Associated with Neurotransmitter Processing

We have identified five new genes by testing for the co-expression of their cDNA with that from genes known to be dopamine- and norepinephrine-related in the libraries comprising LIFESEQ database (Incyte Pharamaceuticals). The five genes show statistically significant

20   associations with the known dopamine- and norepinephrine-related genes as measured by a Fisher exact test. These five new genes are potential therapeutic proteins and therapeutic targets for the treatment of diseases in which the levels of norepinephrine and/or dopamine are perturbed from their normal levels.

The new genes were identified from a total of 41,419 assembled Incyte gene sequences.

25   The degree of association was measured by probability values and has a cutoff of p value less than 0.00001. This was followed by annotation and literature searches to insure that the genes that passed the probability test have strong association with known neurotransmitter-processing-specific genes. This process was reiterated so that the initial 41419 genes were reduced to the final five genes associated with neurotransmitter processing. Details of identification for the five genes are

30   presented in Tables 5 to 10. These tables show the five genes that were most closely associated for each target new gene as measured by coexpression using the Fisher Exact test. The column headings have the same meanings as in Example IV.

Table 5 Coexpression results for Incyte gene 621850

P-value	Associated gene	Occurs	Both occur	Target only	Associated only	Neither
6.91E-09	DBH	12	6	4	6	505

35

8.41E-08	Somatostatin I	29	7	3	22	489
1.37E-07	SCG10	49	8	2	41	470
2.25E-07	Secretogranin II	33	7	3	26	485

Incyte gene 621850 occurred in 10 of 522 cDNA libraries studied and showed strong co-expression with several of the known genes related to neurotransmitter processing, including DBH, somatostatin I and II, as shown in Table 5. 621850 also showed strong association with a SCG10. SCG10 was shown in Example IV to be strongly associated with the known genes related to synthesis and release of dopamine. These results are consistent with the notion that 621850 is associated with neurotransmitter processing. .

Table 6 Coexpression results for Incyte gene 625839

P-value	Associated gene	Occurs	Both occur	Target only	Associated only	Neither
0	DBH	12	7	7	5	502
0	nAchR	13	7	7	6	501
0	TH type 4	15	7	7	8	499
0	Secretogranin II	33	8	6	25	482
0	PARP	34	8	6	26	481

Incyte gene 625839 occurred in 14 of 522 cDNA libraries studied and showed strong co-expression with several of the known genes related to synthesis and release of dopamine, and norepinephrine, including DBH, nAchR, TH type 4, and secretogranin II, as shown in Table 6. 625839 also showed strong association with PARP, a human poly(ADB-ribose) polymerase which is expressed during development in <u>Drosophila</u> (Uchida, K. et al. (1993) 90:3481-3485). These results are consistent with the notion that 625839 is associated with neurotransmitter processing.

Table 7 Coexpression results for Incyte gene 2405140

P-value	Associated gene	Occurs	Both occur	Target only	Associated only	Neither
9.39E-06	DBH	12	3	0	9	509
1.22E-05	nAchR	13	3	0	10	508
1.94E-05	TH type 4	15	3	0	12	506
3.48E-05	AADC	18	3	0	15	503
3.48E-05	Homolog to predicted c. elegans protein F25H2.8	18	3	0	15	503

Incyte gene 2405140 occurred in 3 of 522 cDNA libraries studied and showed strong co-expression with several of the known genes related to neurotransmitter processing, including DBH, nAchR, TH type 4, and AADC, as shown in Table 7. 2405140 also showed strong association with a homolog to a predicted C. elegans protein. F25H2.8. These results are consistent with the notion

that 2405140 is associated with neurotransmitter processing.

Table 8 Coexpression results for Incyte gene 2823339

P-value	Associated gene	Occurs	Both occur	Target only	Associated only	Neither
2.20E-08	Secretogranin I	45	9	3	36	473
4.51E-08	Secretogranin II	33	8	4	25	484
1.37E-07	hVMAT1	8	5	7	3	506
1.56E-07	TH type 4	15	6	6	9	500

Incyte gene 2823339 occurred in 12 of 522 cDNA libraries studied and showed strong co-expression with several of the known genes related to neurotransmitter processing, including secretogranin I and II, hVMAT1 and TH type 4, as shown in Table 8. These results are consistent with the notion that 2823339 is associated with neurotransmitter processing.

Table 9 Coexpression results for Incyte gene 2825861

P-value	Associated gene	Occurs	Both occur	Target only	Associated only	Neither
0	DBH	12	3	0	9	509
0	TH type 4	15	3	0	12	506
0	Human factor X	29	3	0	26	492

Incyte gene 2825861 occurred in 3 of 522 cDNA libraries studied and showed strong co-expression with two of the known genes related to neurotransmitter processing, including DBH and TH type 4, as shown in Table 9. 2825861 also shows association with a blood coagulation factor, human factor X. These results are consistent with the notion that 2825861 is associated with neurotransmitter processing..

VI.   New Genes Associated with Neurotransmitter Processing

Nucleic acids comprising nucleic acid sequences of SEQ ID NOs: 1-5 of the present invention were identified from Incyte Clones 621850, 625839, 2405140, 2823339, and 2825861, respectively, and assembled as described in Example III. The nucleic acid sequences were translated and reading frames determined when possible. A full length protein sequence, SEQ ID NO: 6, was derived from SEQ ID NO: 4 after determining its coding frame. SEQ ID NO: 1-6 were then annotated according to Example VII using BLAST and other motif search tools against databases of known molecules.

SEQ ID NO: 1 showed 99% sequence identity with a human brain-specific mRNA, KIAA0604 (g3043731). In addition, a polypeptide sequence translated from SEQ ID NO:1 showed

about 77.3% sequence homology with the gene that encodes endothelin-converting enzyme 2 of Bos taurus, ECE-2. As described in Example IV, ECEs are known to be involved in the pathway of neurotransmitter processing.

SEQ ID NO: 2 showed 60% sequence identity from about nucleotide 80 to about
5   nucleotide 1818 with a human brain-specific gene, KIAA0604 (g3043731), and 61% sequence identity from about necleotide 144 to about nucleotide 1760 with a human mRNA for endothelin-converting enzyme 1, ECE-1 (g1197803). ECE-1 is a metalloprotease known to be associated with the neurotransmitter processing.

SEQ ID NO:4 showed about 58% sequence identity from nucleotide 332 to nucleotide
10   1393 with the gene which encodes the human heavy neurofilament subunit, NF-H (g35028), a homolog to intermediate filament (IF) proteins. The corresponding amino acid sequence, SEQ ID NO: 6, is 210 amino acids in length and has one potential N-glycosylation site at N16; three potential casein kinase II phosphorylation sites at S21, S100, and S114; one potential glycosaminoglycan attachment site at S110; and two potential protein kinase C phosphorylation
15   sites at S27 and S110.

VII.   Homology Searching for Genes Associated with Neurotransmitter Processing

Polynucleotide sequences, SEQ ID NOs: 1-5, and polypeptide sequences, SEQ ID NO: 6, were queried against databases derived from sources such as GenBank and SwissProt. These databases, which contain previously identified and annotated sequences, were searched for regions
20   of similarity using Basic Local Alignment Search Tool (BLAST; Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410) and Smith-Waterman alignment (Smith, T. et al. (1992) Protein Engineering 5:35-51). BLAST searched for matches and reported only those that satisfied the probability thresholds of 10^{-25} or less for nucleotide sequences and 10^{-8} or less for polypeptide sequences.

The polypeptide sequences were also analyzed for known motif patterns using MOTIFS,
25   SPSCAN, BLIMPS, and Hidden Markov Model (HMM)-based protocols. MOTIFS (Genetics Computer Group, Madison WI) searches polypeptide sequences for patterns that match those defined in the Prosite Dictionary of Protein Sites and Patterns (Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221), and displays the patterns found and their corresponding literature abstracts. SPSCAN (Genetics Computer Group, Madison, WI) searches for potential signal
30   peptide sequences using a weighted matrix method (Nielsen, H. et al. (1997) Prot. Eng. 10:1-6). Hits with a score of 5 or greater were considered. BLIMPS uses a weighted matrix analysis algorithm to search for sequence similarity between the polypeptide sequences and those contained in BLOCKS, a database consisting of short amino acid segments, or blocks, of 3-60 amino acids in length, compiled from the PROSITE database (Henikoff, S. and G. J. Henikoff

(1991) Nucleic Acids Res. 19:6565-6572; Bairoch et al., supra), and those in PRINTS, a protein fingerprint database based on non-redundant sequences obtained from sources such as SwissProt, GenBank, PIR, and NRL-3D (Attwood, T. K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424). For the purposes of the present invention, the BLIMPS searches reported matches

5   with a cutoff score of 1000 or greater and a cutoff probability value of 1.0×10^{-3}. HMM-based protocols were based on a probabilistic approach and searched for consensus primary structures of gene families in the protein sequences (Eddy, S.R. (1996) Cur. Opin. Str. Biol. 6:361-365; Sonnhammer, E.L.L. et al. (1997) Proteins 28:405-420). More than 500 known protein families with cutoff scores ranging from 10 to 50 bits were selected for use in this invention.

10   **VIII.   Labeling and Use of Individual Hybridization Probes**

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech, Piscataway NJ), and T4 polynucleotide kinase (NEN Life Science Products, Boston MA). The labeled oligonucleotides are substantially

15   purified using a SEPHADEX G-25 superfine resin column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (NEN Life Science Products).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to

20   nylon membranes (NYTRAN PLUS, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR film (Eastman Kodak, Rochester NY) is exposed to the blots for several hours, hybridization patterns are compared.

25   **IX.   Production of Specific Antibodies**

SEQ ID NO: 6 substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the amino acid sequence is analyzed using LASERGENE software

30   (DNASTAR Madison WI) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A Peptide synthesizer (PE Biosystems) using Fmoc-chemistry and

coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester to increase immunogenicity. Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA,

5   reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.